

Digital Preservation and Access to Cultural and Scientific Heritage: Presentation of the KT-DigiCult-Bg project

Dr. Nikola IKONOMOV

**Institute for Bulgarian Language, Bulgarian Academy of
Sciences, Sofia, Bulgaria**

nikonov@ibl.bas.bg

Dr. Milena DOBREVA

**Institute of Mathematics and Informatics, Bulgarian
Academy of Sciences, Sofia, Bulgaria**

dobreva@ufal.ms.mff.cuni.cz

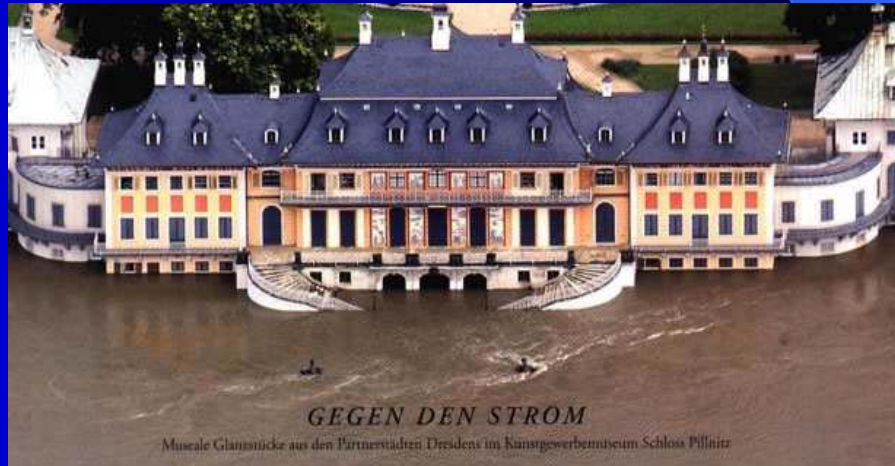
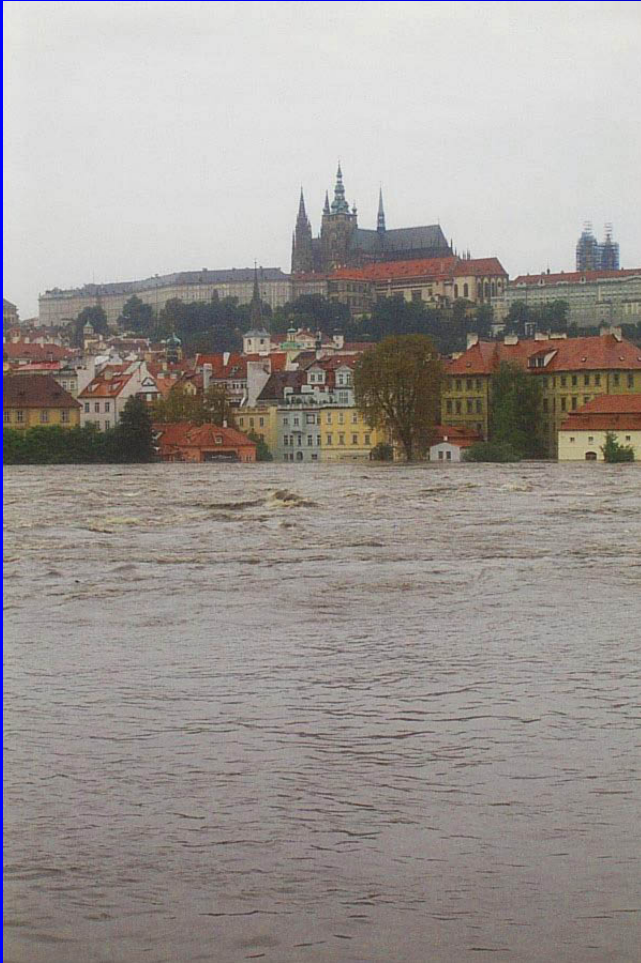
Contents

- Introduction
- Previous Bulgarian experience
- The KT-DigiCult-Bg project
- Conclusions

Introduction

- Digitisation of cultural and scientific heritage is among the EU priorities for two basic reasons:
 - necessity for heritage preservation.
 - new understanding of access to the heritage

Heritage Preservation



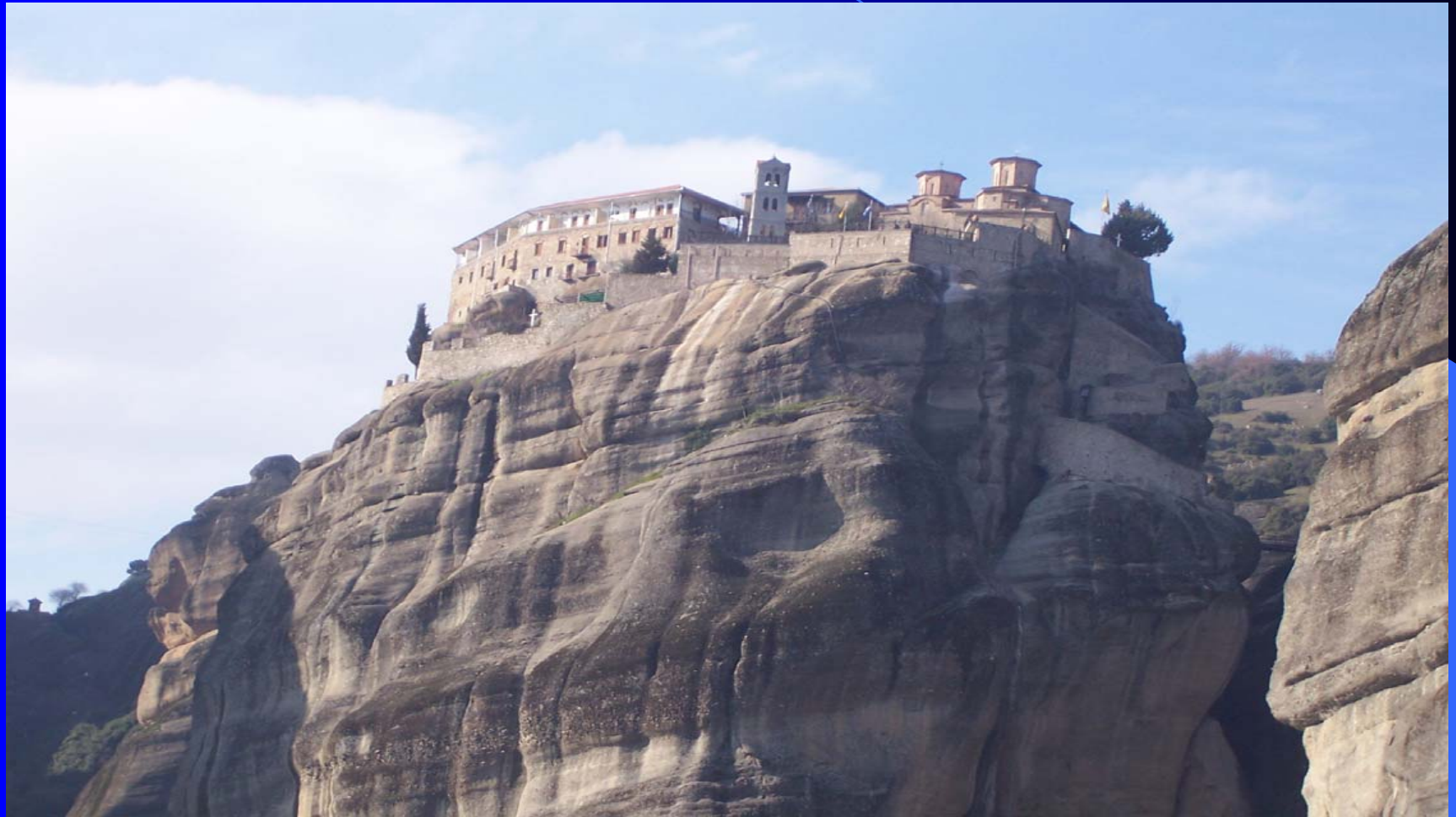
Some ideas about preservation



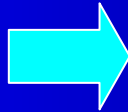
Cultural and scientific heritage preservation

- Preservation means to prepare physical copies of the most valuable products of the human intellect and culture in order to save them for the future generations.

Some aspects of the access to the heritage



From restricted community to the general public

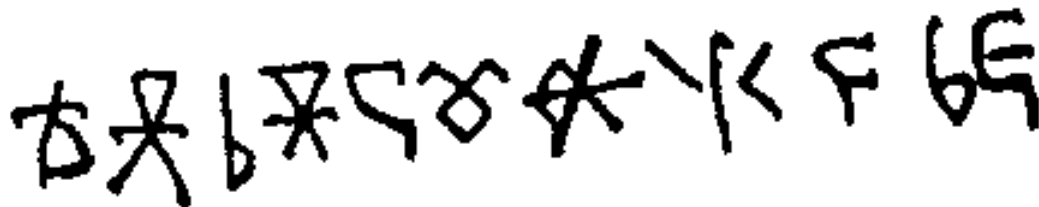




- South-Eastern Europe is underrepresented in the cultural space

Bulgarian heritage assets

- 8500 Christian manuscripts + 4000 Islamic
- More than 35000 early printed books
- Third largest collection of epigraphic inscriptions in Latin and Ancient Greek in the world
- Old Bulgarian runic inscriptions
- ... and numerous other



Old Bulgarian runic inscriptions, showing a sequence of stylized characters.

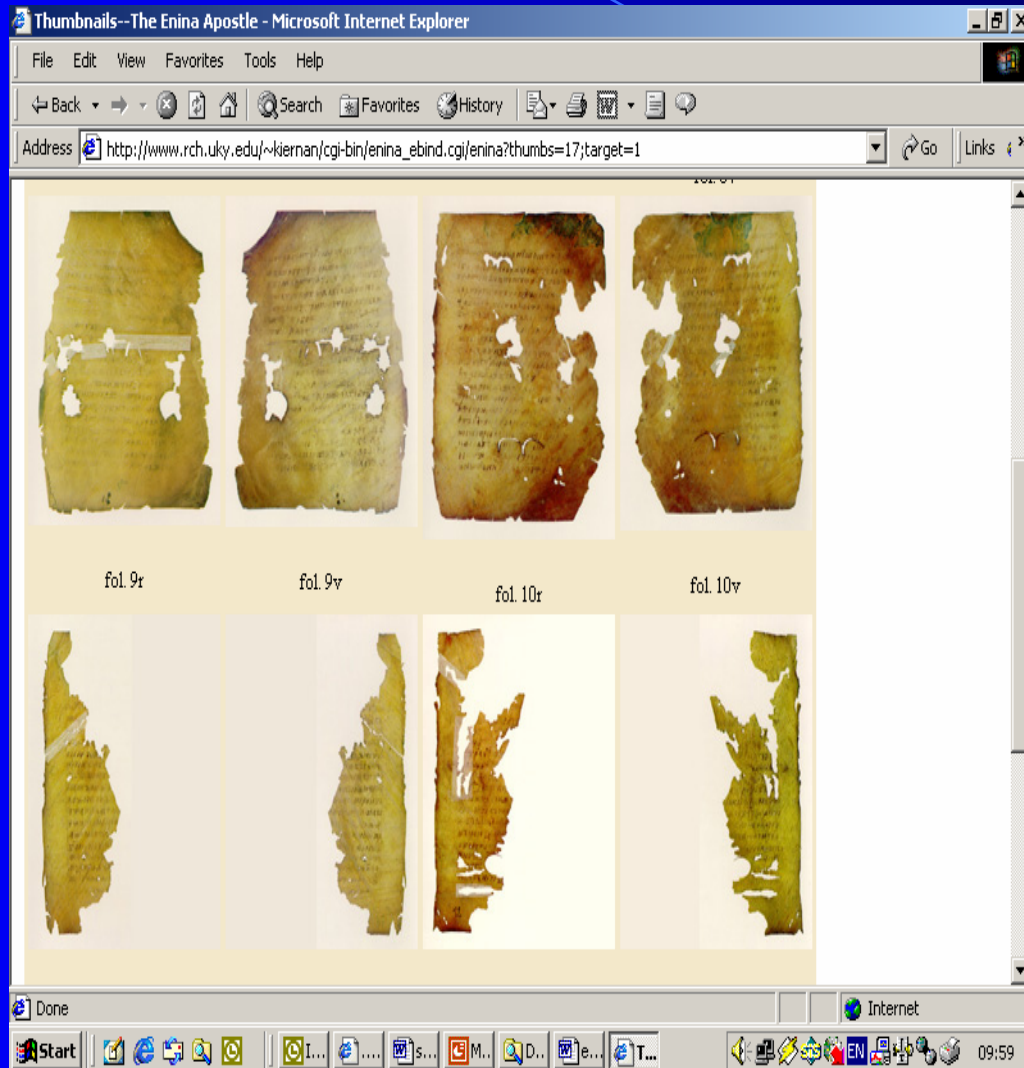
Previous experience – 1

- **Cataloguing:** ISIS (15 years ago), Repertorium (300), MASTER (30), XEditMan (850)
- **Electronic publications on CD-ROMs:** Balkan manuscripts, The Holy Koran, Bulgarian Iconography; The Treasury of Lysimachos

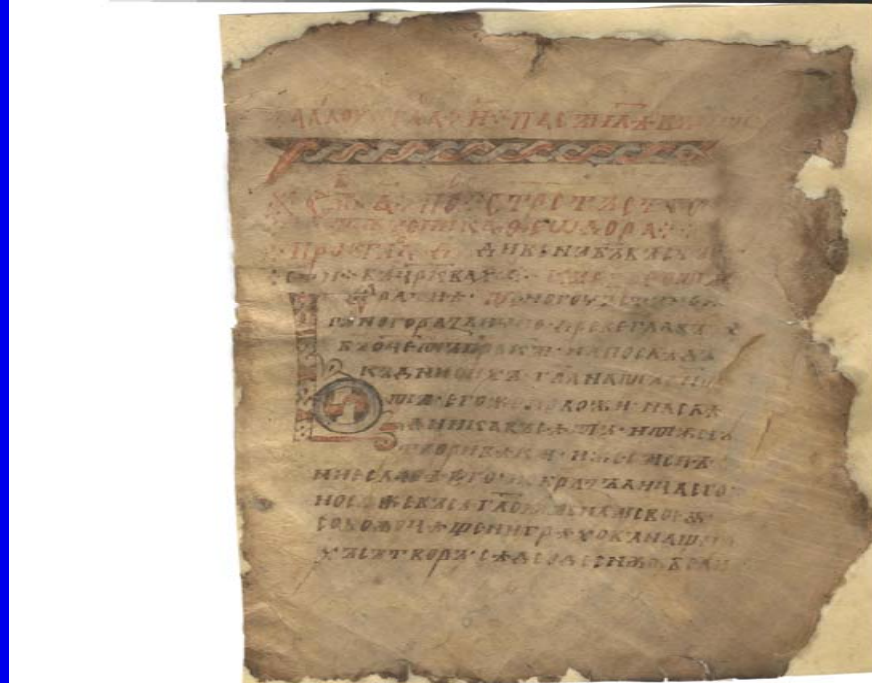
Previous experience – 2

- **Presentation of images** (An Isolated example of digitized images not supplied with texts): The Enina Apostle
 - 11 century (the oldest Bulgarian manuscript stored in Bulgaria)
 - 39 folios survived in a lime-pit, badly damaged

The Enina Apostle



The Enina Apostle



Previous experience - 3



- 3D models of immovable heritage (Bojana church- 13th century)
- Digitization and Restoration of Audio Archives

The KT-DigiCult-Bg project

- The project *Knowledge Transfer for the Digitisation of Cultural and Scientific Heritage in Bulgaria* (KT-DigiCULT-BG) is a four year project supported by the Marie Curie programme.
- It is coordinated by the Institute of Mathematics and Informatics (Bulgaria).
- Project partners are: Det Arnamagnæanske Institut (Københavns Universitet, Denmark), Trinity College (Dublin, Ireland), Charles University (Prague, Czech Republic), and the Institute of Informatics and Telecommunications, National Centre for Scientific Research “Demokritos” (Athens, Greece)

Project specifics

- The project is a typical Transfer of Knowledge action . It is structured as two-phase development scheme:
 - Knowledge acquisition phase
 - Knowledge transfer phase
 - Direct guidance
 - Establishment of a National Digitisation Centre at the Host Institution

Basic field of work

- General methodology and practical setting for digitisation of cultural and scientific heritage.
- Digitisation of mediaeval manuscripts (incl. digital imaging, cataloguing, text representation, electronic publishing).
- Digitisation of mathematical texts and building digital mathematical library of works of Bulgarian mathematicians.
- Virtual reality applications for presentation of immovable cultural heritage.
- Audio archives: methods for digitisation and restoration.
- Application of quantitative methods for the study of data related to the cultural heritage.
- Applications of edutainment to cultural heritage studies.)

Some results after the first project year

- IMI (the host) started the digitisation of its scientific heritage.

АЛГОРИТМ ДЛЯ НАХОЖДЕНИЯ ВСЕХ ИЗОМОРФИЗМОВ ДВУХ ГРАФОВ

АТАНАС А. РАДЕНСКИ

В работе предлагается эффективный алгоритм для нахождения всех изоморфизмов двух неориентированных или ориентированных графов. Время, необходимое для нахождения одного изоморфизма двух n -вершинных графов, пропорционально n^6 . Алгоритм можно использовать, например, для решения некоторых задач химии и теории электрических цепей.

1. Введение. Пусть L — неориентированный связный униграф без петель с вершинами x_1, x_2, \dots, x_n . Элементы матрицы смежности $A_L = (a_{ij})$ графа L определяются следующим образом:

- если вершины x_i и x_j смежны, то $a_{ij} = a_{ji} = 1$;
- если вершины x_i и x_j несмежны, то $a_{ij} = a_{ji} = 0$ ($i, j = 1, 2, \dots, n$)

Отметим, что вершина x_i несмежна с вершиной x_i . Матрица смежности A_L вполне определяет граф L .

Пусть L и M — два графа с матрицами смежности (a_{ij}) и (b_{ij}) и с множествами вершин $X = \{x_1, x_2, \dots, x_n\}$ и $Y = \{y_1, y_2, \dots, y_n\}$. Графы L и M называются изоморфными ($L \cong M$), если существует взаимно однозначное отображение φ множества X на множество Y , такое, что из равенств $y_p = \varphi(x_i)$, $y_q = \varphi(x_j)$ следует $a_{ij} = b_{pq}$ ($i, j = 1, 2, \dots, n$). Теоретически проблема нахождения всех изоморфизмов графов L и M решается просто — проверкой всех $n!$ взаимно однозначных отображений φ множества X на Y . Однако практически такое решение неприменимо. Для некоторых специальных графов существуют эффективные решения проблемы изоморфизма [2, 3, 4]. В работе [1] предлагается общий алгоритм для нахождения всех изоморфизмов двух графов, у которого порядок быстродействия — n^6 . Этот алгоритм основан на нахождении длинного списка характеристик. Совокупность списков предложенного вида оказывается инвариантной относительно изоморфизма. В [6] описан один из интереснейших алгоритмов для проверки изоморфности двух графов, в котором используется одно недоказанное предположение.

В настоящей работе предлагается весьма простая характеристика, определяющая граф с точностью до изоморфизма. Эта характеристика позволяет выяснить, изоморфен ли граф L графу M , и если они изоморфны, найти все изоморфизмы. Предложенный алгоритм без особых изменений применим и к случаю, когда L и M — произвольные униграфы.

2. Определения. Подграфом графа L называется каждый граф L' с множеством вершин $X', X' \subseteq X$, матрица смежности которого получена из A_L путем удаления всех строк и столбцов с номерами i , для которых $x_i \notin X'$ и $x_i \notin X'$ ($i = 1, 2, \dots, n$), т. е. две вершины из X' смежны

СЕРДИКА. Българско математическо списание. Том 1, 1975, с. 167—177.

138

ИЗВЕСТИЯ ИМ. ВУЛКА

Следствие 1. Пусть $Y = (y, \infty)$ -компактное подпространство вполне регулярного пространства X . Тогда для любой теоретико-множественной τ -операции ψ и любого $n < \aleph^+$ имеем:

$$\psi - \bar{\psi}_n(Y) = \{Y \cap L \mid L \in \psi - \bar{\psi}_n(X)\},$$

$$\psi - U_n(Y) = \{Y \cap L \mid L \in \psi - U_n(X)\},$$

$$\psi - \bar{\psi}_n^*(Y) = \{Y \cap L \mid L \in \psi - \bar{\psi}_n^*(X)\}.$$

Следствие 2. Пусть $Y = (y, \infty)$ -компактное подпространство вполне регулярного пространства X . Если для некоторой теоретико-множественной τ -операции ψ пространство Y удовлетворяет свойству $\psi - H$, то и X удовлетворяет свойству $\psi - H$.

Теорема 4. Пусть $f: X \rightarrow Y$ — совершенное отображение вполне регулярного пространства X на вполне регулярное пространство Y веса τ . Если для некоторой свободной топологии $\mathfrak{S}(m)$ множества Y τ -операция ψ пространства Y удовлетворяет свойству $\psi - H$, то и пространство X удовлетворяет свойству $\psi - H$.

Полнотелостью этого факта можно пользоваться при построении борзовских сетей для многозначных отображений.

3. Модификация топологии. Пусть \mathfrak{S} — топология на множестве X . Семейство $\mathfrak{M} = \{U_\alpha\}$ образует базис некоторой новой топологии $\mathfrak{S}(m)$ на множестве X . Топологию $\mathfrak{S}(m)$ назовем m -модификацией топологии \mathfrak{S} .

Предложение 1. Пусть \mathfrak{S} — класс τ -операций, \mathfrak{S}_1 и \mathfrak{S}_2 семейства подмножеств множества Z и $\mathfrak{S}_3 = \psi(\mathfrak{S})$ для некоторой операции $\psi \in \mathfrak{S}$. Если для любой операции $\psi \in \mathfrak{S}$ семейство $\psi(\mathfrak{S}_1)$ не инвариантно относительно дополнения, то и семейство $\psi(\mathfrak{S}_2)$ не инвариантно относительно дополнения для каждой операции $\psi \in \mathfrak{S}$.

Лемма 1. Пусть $\psi \in \mathfrak{S}$. Положим $\psi^* = \psi(\mathfrak{S})$. Тогда $\psi(\mathfrak{S}_1^*) = \psi(\mathfrak{S}_2^*) = \psi^*$. Поскольку ψ^* — семейство $\psi(\mathfrak{S})$ не инвариантно относительно дополнения. Предположение доказано.

Положим $\mathfrak{M} = \{X - \mathfrak{S}(m) - P\}$, где $P \in \mathfrak{S}$. Если (X, \mathfrak{S}) — топологическое пространство веса τ и $2^m \leq \tau$, то $\mathfrak{S}(m) = \psi_m(\mathfrak{S})$.

Следствие 3. Пусть (X, \mathfrak{S}) — вполне регулярное пространство веса τ и $2^m \leq \tau$. Если для любой теоретико-множественной τ -операции ψ пространство $(X, \mathfrak{S}(m))$ удовлетворяет свойству $\psi - C$, то и пространство $(X, \mathfrak{S}(m))$ удовлетворяет свойству $\psi - C$ для любой теоретико-множественной τ -операции ψ .

В множестве J рассмотрим семейство подмножеств N_i , где:

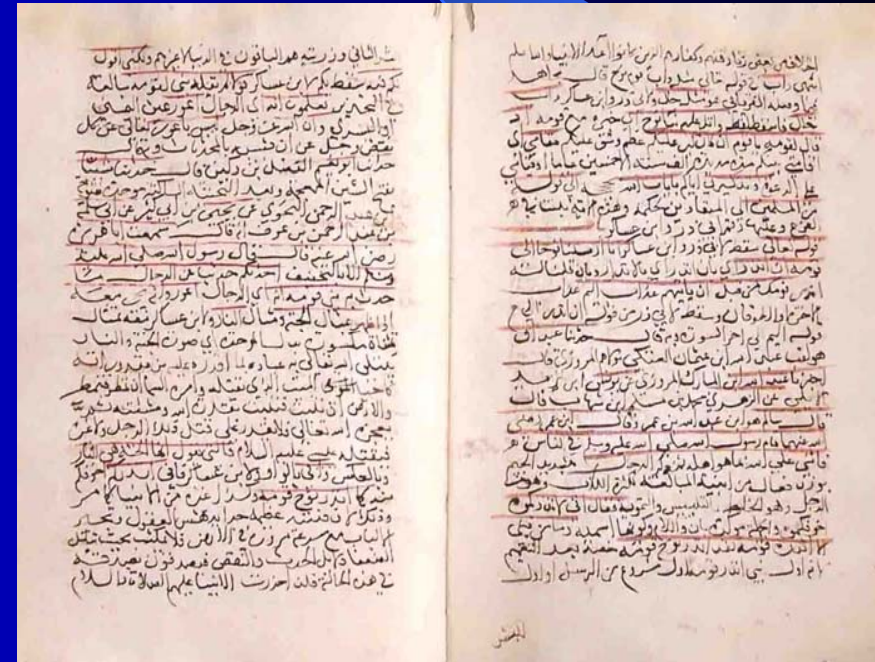
- 1) если $y \in N_i$, то $|x| \leq \tau$;
- 2) если $y_1, y_2 \in N_i$ и $y_1 \neq y_2$, то $y_1 \cap y_2 = \emptyset$;
- 3) $y \in N_i$;
- 4) если $m < \tau$, то в N_i имеется ровно τ различных элементов мощности m .

Теперь положим $K = \{y \subseteq J, y \subseteq m \text{ для некоторого } y \in N_i\}$. Положим $\sigma = \psi(K)$. Операция ψ является положительной*. Пусть P_i является операцией пересечения менее τ элементов. Тогда $\psi_i = \psi - S(P_i)$.

* — операция ψ называется положительной, если для любых множеств $\{E_1, \dots, E_n\}$, $\{A_1, \dots, A_n\}$ из $E_1 \times \dots \times E_n$ имеем $\psi(\{A_i\}) \subseteq \psi(\{E_i\})$. Положительные теоретико-множественные τ -операции называются τ - σ -операциями.

Some results after the first project year

- A CD- “Sofia. Religious spaces” was prepared and published



Some results after the first project year

- An Old Cyrillic UNICODE font based on Codex Suprasliensis script (11 c.) was designed

ПРѢЖДЕУЕТЫРЬКАЛАНДЪМАРТА. СНОРѢУЪ

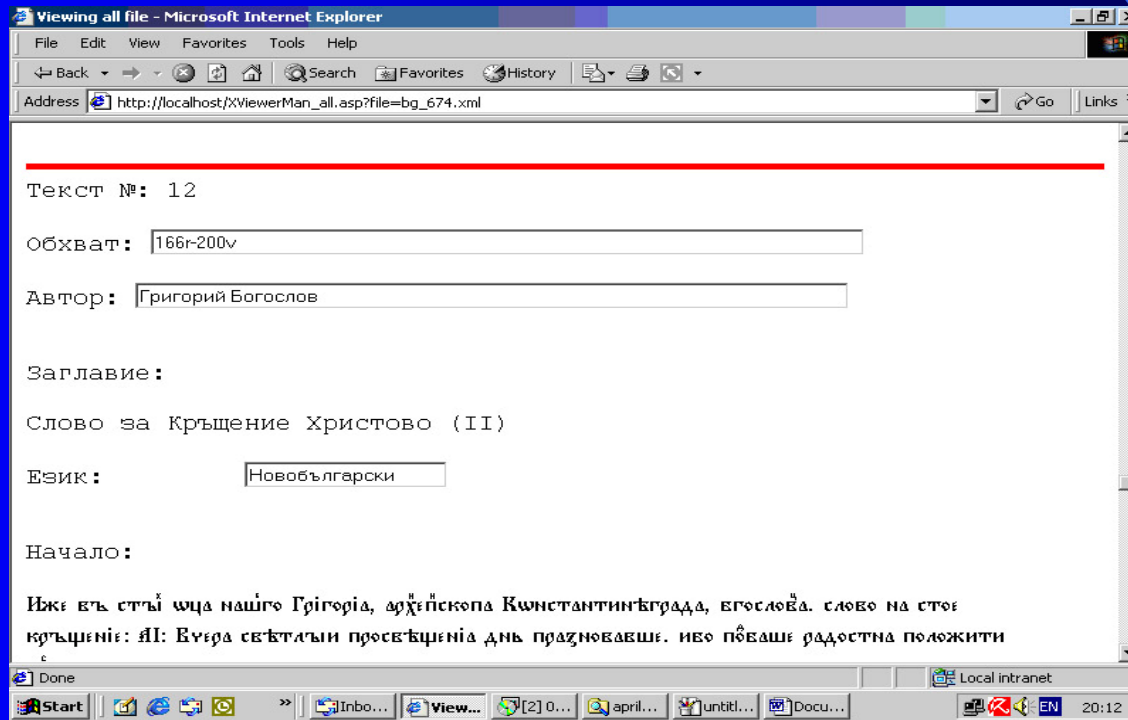
ПРѢЖДЕУЕТЪИРКАЛАНДЪМАРТА. СНОРѢУЪ
ВЪКЪСФЕУРОАРА. ПРѢДАШАЖЕСВОАДОУША
ГОСПОДЕВН. ПРѢЖДЕУДАНЪМАРТА. ПРНЛН
КННННСАМОВАЛСТЪНН. НАЖЕЖЕУЪСА РЪСГВХ
ЖШТОУГОУНАШЕМОУНБѢУНЕСПОУ. Н ВЛА
ДЪИЧНАШЕМОУИСХСОУ. КМОУЖЕНЕСТЪСЛА
ВАНДРЪЖАВАНУЕСТЪ. НЪННННПРНМОИВЪ
ВЪКЪКЪКОМЪ. АМННЪ.



СТААГОВАСНАЪ АРХНЕПЦАКЕСАРНАКАПАДОСНИКЪА.

Some results after the first project year

- A specialized XML Editor for Manuscript Data for editing and browsing catalogue descriptions of mediaeval manuscripts was developed



Some results after the first project year

- a methodology of using wavelets for image enhancement and character recognition was elaborated and implemented.

Conclusion

- Current priorities of EU: IST 2.5.10 (Access to and preservation of cultural and scientific resources)
 - Enriched conceptual representations
 - Advanced access methods
 - Long-term preservation
- We follow the recent developments but still have to do mass digitisation

Conclusions

- The rich heritage of Bulgaria is still expecting to take its legitimate place in the world cultural space. We hope that through the project we will raise the interest of the scientific and cultural heritage institutions in the country and will push them to start real work.
- The project gives opportunities to broaden the international cooperation in the field of digitisation of cultural and scientific heritage.

Conclusions

- We highly appreciate the experience of other small countries in Europe, and in particular of institutions from the Czech Republic. The similarity of territory and population, the richness of the cultural and scientific heritage, and the common historical fate in the last century rise comparable problems. The way they are or have been solved in the Czech Republic and the visible progress in the field give promises for a successful Bulgarian digitisation.

Hope I am in time???



Thank you!!!